

**Comments to the**  
**United States Office of Science and Technology Policy (OSTP)**  
**on**  
**An Artificial Intelligence (AI) Action Plan**  
**Center for AI and Digital Policy**

The Center for AI and Digital Policy submits these Comment in response to the Request for Information (RFI) issued by the Office of Science and Technology Policy (OSTP) on the Development of an Artificial Intelligence (AI) Action Plan.<sup>1</sup> The RFI states that the AI Action Plan will be developed pursuant to President’s January 2025 Executive Order.<sup>2</sup> The purpose of that Executive Order is to “solidify our position as the global leader in AI and secure a brighter future for all Americans.”<sup>3</sup>

U.S. federal AI policy has advanced largely through executive action upon the advice and guidance of OSTP. In 2019 and 2020, President Trump issued two Executive Orders on AI.<sup>4</sup> Several of the provisions contained in these Executive Orders were carried forward in the 2023 Executive Order on Safe, Secure and Trustworthy AI.<sup>5</sup> The OMB, under the first Trump Administration, issued guidance on the deployment of AI systems across the federal

---

<sup>1</sup> OSTP, National Science Foundation, *Request for Information on the Development of an Artificial Intelligence (AI) Action Plan*, 90 FR 9088, 2025-02305, Feb. 6, 2025 - [Federal Register :: Request for Information on the Development of an Artificial Intelligence \(AI\) Action Plan](#)

<sup>2</sup> Executive Office of the President, *Removing Barriers to American Leadership in Artificial Intelligence*, EO 14179, 90 FR 8741, 2025-02172, Jan. 23, 2025, <https://www.federalregister.gov/documents/2025/01/31/2025-02172/removing-barriers-to-american-leadership-in-artificial-intelligence> (“The January 2025 Executive Order”)

<sup>3</sup> Section 1, EO 14179

<sup>4</sup> President Donald Trump, *Executive Order 13960 - Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government*, 85 FR 78939, Dec. 3, 2020, <https://www.federalregister.gov/documents/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal-government> [“EO 13960”]; President Donald Trump, *Executive Order 13859- Maintaining American Leadership in Artificial Intelligence*, 84 FR 3967, Feb. 11, 2019, <https://www.federalregister.gov/documents/2019/02/14/2019-02544/maintaining-american-leadership-in-artificial-intelligence> [“EO 13859”]

<sup>5</sup> Executive Order 14110, *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, Federal Register Vol. 88, No. 210, Oct. 30, 2023, <https://www.govinfo.gov/content/pkg/FR-2023-11-01/pdf/2023-24283.pdf> [“The 2023 Executive ”]

government.<sup>6</sup> Many of these recommendations were carried forward by the OMB during the Biden Administration.<sup>7</sup>

The January 2025 Executive Order on AI states, “It is the policy of the United States to sustain and enhance America’s global AI dominance in order to promote human flourishing, economic competitiveness, and national security.”<sup>8</sup> The Executive Order Fact Sheet states, “Today’s Executive Order builds upon these past successes and clears a path for the United States to act decisively to retain leadership in AI, rooted in free speech and human flourishing.”<sup>9</sup>

## About CAIDP

The Center for AI and Digital Policy (CAIDP) is an independent, non-profit research and education organization that advises national governments and international organizations on artificial intelligence (AI) and digital policy.<sup>10</sup> CAIDP supports AI policies that advance the American values such as privacy, civil liberties, and civil rights, and promotes broad social inclusion based on fundamental rights, democratic institutions, and the rule of law.

## CAIDP Recommendations for the US AI Action Plan

We welcome the RFI initiated by the OSTP and the commitment to “incorporating the public’s comments and innovative ideas.”<sup>11</sup> Our key recommendations to promote human flourishing, American competitiveness, and a brighter future for all Americans are as follows:

---

<sup>6</sup> Russell T. Vought, Office of Management and Budget (OMB), *Guidance for Regulation of Artificial Intelligence Applications*, M-21-06, Nov. 17, 2020, <https://trumpwhitehouse.archives.gov/wp-content/uploads/2020/11/M-21-06.pdf> [“M-21-06”]

<sup>7</sup> Shalanda D. Young, Office of the Management and Budget, *Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence*, M-24-10, Mar. 28, 2024, <https://bidenwhitehouse.archives.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf> (“M-24-10”); Shalanda D. Young, Office of the Management and Budget, *Advancing the Responsible Acquisition of Artificial Intelligence in Government*, M-24-18, Sept. 24, 2024, <https://bidenwhitehouse.archives.gov/wp-content/uploads/2024/10/M-24-18-AI-Acquisition-Memorandum.pdf> (“M-24-18”)

<sup>8</sup> EO 14179, Sec. 2

<sup>9</sup> The White House, *FACT SHEET: PRESIDENT DONALD J. TRUMP TAKES ACTION TO ENHANCE AMERICA’S AI LEADERSHIP* (Jan. 23, 2025), <https://www.whitehouse.gov/fact-sheets/2025/01/fact-sheet-president-donald-j-trump-takes-action-to-enhance-americas-ai-leadership/>

<sup>10</sup> CAIDP, <https://www.caidp.org>

<sup>11</sup> The White House, *Public Comment Invited on Artificial Intelligence Action Plan*, Briefings & Statements, (Quoting comments of Lynne Parker, Principal Deputy Director of the Office of Science and Technology Policy), Feb. 25, 2025, <https://www.whitehouse.gov/briefings-statements/2025/02/public-comment-invited-on-artificial-intelligence-action-plan/>

## 1) AI innovation

The US AI Action Plan should increase federal investment in AI research and education and encourage the development of innovative AI systems that are less energy-dependent and less data-intensive. Encourage the diffusion of AI innovation across the board rather than restricting it to a few corporations. The 2020 OSTP American AI Initiative Report highlighted innovation within the federal government exploring high-performance, energy-efficient hardware and machine-learning architectures and enabling researchers they support.<sup>12</sup> The infrastructure, cost, and current compute requirements is one of the reasons smaller companies cannot easily enter the AI sector.<sup>13</sup> Setting the policy and regulatory incentives for less energy intensive data systems can promote competition and innovation in the AI industry aligning with the objectives of the January 2025 Executive Order.

The OSTP should also promote the use of Privacy Enhancing Technologies (PETs) to minimize or eliminate the collection of personal data to encourage AI innovation safely.<sup>14</sup> The OSTP has long recognized that “PETs include utilizing low-data artificial intelligence, deleting unnecessary data, and creating techniques for robust anonymity.”<sup>15</sup> By encouraging investment in AI that is built on PETs and in turn building capacity for independent evaluation of PETs, the OSTP can move American AI to an ecosystem that expands horizontal and vertical innovation, improves resilience to hacking and espionage, develops capability, and creates new jobs due to workforce requirements.

## 2) American Leadership

The US AI Action Plan should maintain American leadership in global AI Policy to promote human flourishing, economic competitiveness, and national security and ratify the US-sponsored AI Convention for freedom, human rights, and the rule of law.<sup>16</sup> The First Trump Administration was instrumental in advancing US AI policy leadership when the US led other OECD countries to draft and endorse the OECD AI Principles, an influential framework for AI governance.<sup>17</sup> Michael Kratsios, then US Chief Technology Officer, stated “Our vision for

---

<sup>12</sup> The Office of Science and Technology Policy (OSTP), *American Artificial Intelligence Initiative: Year One Annual Report*, pg. 11 (Feb. 2020), <https://www.nitrd.gov/nitrdgroups/images/c/c1/American-AI-Initiative-One-Year-Annual-Report.pdf> [“American AI Initiative Report”]

<sup>13</sup> CAIDP, *Comments to the Autorité de la concurrence on the generative artificial intelligence sector*, Mar. 22, 2024,

<sup>14</sup> CAIDP, *Comments to OSTP on the Promotion of Privacy Enhancing Technologies (PETs)*, (Jul.8, 2022), <https://www.caidp.org/app/download/8402029763/CAIDP-PETS-OSTP-07082022.pdf>

<sup>15</sup> *Id.*

<sup>16</sup> Council of Europe, *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law* (2024), <https://rm.coe.int/1680afae3c7>

<sup>17</sup> Trump White House, *Artificial Intelligence for the American People*, International Leadership on AI, <https://trumpwhitehouse.archives.gov/ai/ai-american-values/>

artificial intelligence is rooted in the rule of law, respect for rights, and the spirit of freedom.”<sup>18</sup> We urge the OSTP to advise President Trump to move forward with the ratification of the AI Convention to protect human rights, democracy and the rule of law.

### 3) AI Safety Standards

The US AI Action Plan should fully fund the AI Safety Institute and ensure US leadership in international standards for the safe deployment of AI systems. The 2020 OSTP American AI Initiative Report stated, “The National Institute for Standards and Technology is a leader in foundational research that informs the development of technical standards for reliable, robust, and trustworthy AI.”<sup>19</sup> The AI Safety Institute created within NIST fulfils an important function of ensuring American AI is safe, resilient, and ahead of emerging risks of advanced AI models.

The OSTP 2019 National AI R&D Strategic Plan<sup>20</sup> set out eight priority areas including ensuring the safety and security of AI systems; developing shared public datasets and environments for AI training and testing; measuring and evaluating AI technologies through standards and benchmarks. This was carried forward by the following administration which promoted the development and implementation of repeatable processes and mechanisms to understand and mitigate risks related to AI adoption, including with respect to biosecurity, cybersecurity, national security, and critical infrastructure.<sup>21</sup>

The AISI is focused on preventing misuse of advanced AI systems to secure public safety and national security.<sup>22</sup> The US and UK AI Safety Institutes recently published a technical report detailing a pre-deployment evaluation of the upgraded version of Claude 3.5 Sonnet.<sup>23</sup> It addresses the evidence gap concerning the validity, reliability, and practicality of existing general-purpose AI risk assessment methods and develops measurement science – a priority outlined by OSTP

---

<sup>18</sup> Daniel Castro, *Remarks by Michael Kratsios, U.S. CTO at Center for Data Innovation Forum on AI* Center for Data Innovation (Sept. 18, 2019), <https://datainnovation.org/2019/09/remarks-by-michael-kratsios-u-s-cto-at-center-for-data-innovation-forum-on-ai/>

<sup>19</sup> American AI Initiative Report, pg. 7

<sup>20</sup> OSTP, *The National AI R&D Strategic Plan: 2019 Update* (June 2019), <https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf>

<sup>21</sup> Congressional Research Service, *Highlights of the 2023 Executive Order on Artificial Intelligence for Congress*, 2025, <https://www.congress.gov/crs-product/R47843>.

<sup>22</sup> U.S. Artificial Intelligence Safety Institute, <https://www.nist.gov/aisi>

<sup>23</sup> Paris AI Action Summit, *International AI Safety Report*, p. 185 (Jan. 2025), [https://assets.publishing.service.gov.uk/media/679a0c48a77d250007d313ee/International\\_AI\\_Safety\\_Report\\_2025\\_accessible\\_f.pdf](https://assets.publishing.service.gov.uk/media/679a0c48a77d250007d313ee/International_AI_Safety_Report_2025_accessible_f.pdf)

Director, Michael Kratsios in his Senate confirmation testimony<sup>24</sup> and in public statements on his vision for AI policy as OSTP Director.<sup>25</sup>

As the speed and scale of AI technologies expand both domestically and globally, much care must be given to ensure the proper recruitment and retention of AI researchers and practitioners to ensure a viable pool of talent for tomorrow's intellectual demands.<sup>26</sup> The OSTP should prioritize fully funding and building on AISI's mission to develop critical capabilities to lead in advance AI evaluations, standard-setting, and threat assessment.

#### 4) Procurement

The US AI Action Plan should build on the principles set out in Section 3 of the 2020 Executive Order *Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government*.<sup>27</sup> CAIDP endorses all of these principles in the 2020 Executive Order for AI Systems: (a) **Lawful and respectful of our Nation's values**; (b) **Purposeful and performance-driven**; (c) **Accurate, reliable, and effective**; (d) **Safe, secure, and resilient**; (e) **Understandable**; (f) **Responsible and traceable**; (g) **Regularly monitored**; (h) **Transparent**, and (i) **Accountable**.

Successive administrations have focused on increasing federal AI capacity to increase the efficiency and speed of transactions and lower costs.<sup>28</sup> As CAIDP President Merve Hickok explained, "AI expands our research capabilities, gives us greater ability to prototype and simulate and hence innovate faster; helps us better analyze natural, economic, and social patterns; allows services and products to reach wider parts of the population, in more effective ways. However, AI developed and deployed without safeguards poses real challenges."<sup>29</sup>

---

<sup>24</sup> United States Senate Commerce Committee, *Nomination Hearing for Michael Kratsios to lead the Office of Science & Technology Policy and Mark Meador to Serve as Federal Trade Commissioner*, Feb. 25, 2025, <https://www.commerce.senate.gov/2025/2/nominations-hearing-for-michael-kratsios-to-lead-the-office-of-science-and-technology-policy-and-mark-meador-to-serve-as-a-federal-trade-commissioner>

<sup>25</sup> Oxford Generative AI Summit 2024, *Fireside chat: Michael Kratsios, Managing Director of Scale AI and former Chief Technology Officer of the United States* (Dec. 13, 2024), <https://www.youtube.com/watch?v=ODw372xkZh0>

Moderated by: Dr. Keegan McBride, Lecturer in AI, Government, and Policy at the Oxford Internet Institute

<sup>26</sup> CAIDP, *Comments to OSTP on the National Artificial Intelligence Research and Development Strategic Plan*, pg. 11 (Mar. 4, 2022), <https://www.nitrd.gov/rfi/ai/2022/87-FR-5876/NAIRDSP-RFI-2022-combined.pdf>

<sup>27</sup> EO 13960

<sup>28</sup> Merve Hickok, *Public procurement of artificial intelligence systems: new risks and future proofing*, AI & Society, 39, 1213–1227, 2024, <https://doi.org/10.1007/s00146-022-01572-2>

<sup>29</sup> Merve Hickok, *Testimony and Statement for the Record*, "Advances in AI: Are We Ready For a Tech Revolution?" House Committee on Oversight and Accountability, Subcommittee on Cybersecurity, Information Technology, and Government Innovation, (Mar. 8, 2023),



The United States has the highest investment in R&D and public procurement and should demand algorithmic accountability from those benefiting from the R&D funds and contracts.<sup>30</sup> Agencies should be able to switch vendors if and when needed. Such need may arise from vendor's performance, emerging risks due to new AI use cases or cybersecurity, or cost reasons.<sup>31</sup> "The US federal government is the largest buyer in the world and seen as a role model by other countries."<sup>32</sup> With regard to procurement, CAIDP's recommendations are threefold:

- 1) The AI Use Case Inventory process established under EO 13960 should be continued and expanded.<sup>33</sup>
- 2) The minimum risk management practices for government use of AI that impact people's rights or safety set out in OMB Memo M-24-10<sup>34</sup> should be continued
- 3) Sections 4 and 5 of the OMB AI Acquisition Guidance<sup>35</sup> relating to managing risks and performance and ensuring a competitive marketplace should be continued

## 5) Human-centric AI

The US AI Action Plan should ensure that human roles and responsibilities are clearly defined, and appropriately assigned for the design, development, acquisition, and use of AI. Build human capacity and expertise in AI governance to enable policymakers and practitioners to make informed decisions. "Human-centric" is a core goal for safe, secure, and trustworthy AI. This includes appreciating the role of users in design processes, followed by the identification and involvement of additional participants in the design, development, and deployment of AI

---

[https://oversight.house.gov/wp-content/uploads/2023/03/Merve-Hickok\\_testimony\\_March-8th-2023.pdf](https://oversight.house.gov/wp-content/uploads/2023/03/Merve-Hickok_testimony_March-8th-2023.pdf); CAIDP, *Comment to Office of Management & Budget - Responsible Procurement of Artificial Intelligence in Government* (Apr. 29, 2024), [https://www.linkedin.com/posts/center-for-ai-and-digital-policy\\_caiddp-comments-omb-procurement-apr-29-activity-7191112979478659072-Das6](https://www.linkedin.com/posts/center-for-ai-and-digital-policy_caiddp-comments-omb-procurement-apr-29-activity-7191112979478659072-Das6)

<sup>30</sup> *Id.*; CAIDP, *Letter to Shalanda Young, Director, Office of Management & Budget* (Apr. 24, 2023), <https://www.caiddp.org/app/download/8454950563/CAIDP-Statement-OMB-04242023.pdf>

<sup>31</sup> Merve Hickok, *Response to Request for Information - Responsible Procurement of AI in Government*, (Apr. 29, 2024),

<sup>32</sup> Merve Hickok, Evanna Hu, *Don't Let Governments Buy AI Systems That Ignore Human Rights*, *Issues in Science and Technology* 40, no. 3, Spring 2024, 37–41. <https://doi.org/10.58875/ACOG7116>

<sup>33</sup> Executive Office of the President, *Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government*, Executive Order 13960, 85 FR 78939, 2020-27065, <https://www.federalregister.gov/documents/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal-government>

<sup>34</sup> Office of Management and Budget, *Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence*, M-24-10 (Mar. 28, 2024), <https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf>;

<sup>35</sup> Office of Management & Budget, *Advancing the Responsible Acquisition of Artificial Intelligence in Government*, M-24-18 (Sept. 24, 2024), <https://www.whitehouse.gov/wp-content/uploads/2024/10/M-24-18-AI-Acquisition-Memorandum.pdf>

systems.<sup>36</sup> Most importantly, it requires assigning responsibility and accountability for the operation and outcomes of an AI system to an identifiable individual or organization. Truly autonomous AI systems should not be deployed.

## 6) Public Safety

The US AI Action Plan should protect the safety of Americans and ensure that dangerous AI systems, such as lethal autonomous weapons and mass surveillance, are not deployed by the federal government or the private sector. Establish red lines prohibiting AI systems that violate Americans rights such as privacy, civil liberties, and civil rights.<sup>37</sup> A recent report of leading AI experts identifies red lines for AI systems, including limits on unacceptable use and unacceptable behavior of advanced AI systems.<sup>38</sup> Establishing clear limits on unacceptable AI is the foundation for building provably safe and beneficial AI systems.<sup>39</sup>

AI services will turbocharge risks to national security, public safety, online child safety, cybersecurity, deception, and election integrity.<sup>40</sup> Advanced AI systems pose systemic risks including risks to critical cyber infrastructure, payments infrastructure, large-scale harmful content.<sup>41</sup> There are also more subtle ways in which an AI technology could end up with a larger scale of impact than its creators anticipated.<sup>42</sup>

To ensure America's national security and public safety interests OSTP should implement measures such that the federal government oversees safety protocols for advanced AI design and development, ensures rigorous evaluation through NIST/AISI, require incident reporting,<sup>43</sup> and

---

<sup>36</sup> *Maintaining Control Over AI*, Issues in Science and Technology 37, no. 3, Spring 2021, <https://issues.org/debating-human-control-over-artificial-intelligence-forum-shneiderman/>

<sup>37</sup> Christabel Randolph and Marc Rotenberg, *The AI Red Line Challenge* (September 3, 2024) <https://www.techpolicy.press/the-ai-red-line-challenge/>

<sup>38</sup> World Economic Forum, *AI red lines: the opportunities and challenges of setting limits* (Mar. 11, 2025), <https://www.weforum.org/stories/2025/03/ai-red-lines-uses-behaviours/>

<sup>39</sup> *Id.*

<sup>40</sup> CAIDP, *Complaint to FTC – in re OpenAI and ChatGPT* (Mar. 30, 2023), <https://www.caidp.org/cases/openai/>; Open AI, *The GPT-4 System Card* (Mar. 15, 2023), <https://cdn.openai.com/papers/gpt-4-system-card.pdf>; CAIDP, *ChatGPT and the Federal Trade Commission: Still No Guardrail*, (Jul. 9, 2024), [https://www.linkedin.com/posts/center-for-ai-and-digital-policy\\_caidp-still-no-guardrails-july-2024-activity-7216406912035106816-MUCs](https://www.linkedin.com/posts/center-for-ai-and-digital-policy_caidp-still-no-guardrails-july-2024-activity-7216406912035106816-MUCs)

<sup>41</sup> Bengio et al., *Managing AI Risks in an Era of Rapid Progress*, arXiv (2023), <https://managing-ai-risks.com/>; CAIDP, *Complaint to the Federal Trade Commission - In Re OpenAI*, para.71-75, pg. 17,18 (Mar. 30, 2023) , <https://www.caidp.org/cases/openai/>

<sup>42</sup> Stuart Russell, Andrew Critch, *TASRA: a Taxonomy and Analysis of Societal-Scale Risks from AI* (Jun. 14, 2023), <https://arxiv.org/abs/2306.06924>

<sup>43</sup> OECD *AI Incidents Monitor* (AIM), <https://oecd.ai/en/incidents>

oversight by independent experts.<sup>44</sup> These measures will ensure that American AI is safe, effective, competitive, and trustworthy. It can also protect against geopolitical threats, attacks, and adversarial actions.

Insiders from within the AI companies have warned of the information asymmetries and unknown capabilities of advanced AI systems.<sup>45</sup> MIT led an open call by 350+ AI, legal, and policy experts for “A Safe Harbor for Independent AI Evaluation” citing concerns over current practices and policies of AI companies that can chill independent evaluation.<sup>46</sup> OSTP should initiate interagency collaboration to offer whistleblower protections to those who identify and flag serious risks.

## 7) Rights Protections

The US AI Action Plan should protect the rights of Americans and ensure that AI systems that violate American values such as privacy, civil liberties, and civil rights, are not deployed by federal agencies or the private sector. Too many AI systems are designed around spurious concepts and correlations. Many prevalent approaches, including biometric categorization, biometric emotion analysis, and predictive policing, lack scientific validity. Though vendors claim their systems are objective and capable of predicting or inferring such things as a person’s emotions, ethnicity, or likelihood of committing a crime, consensus is growing that they instead identify spurious correlations between disparate data points.<sup>47</sup>

The OSTP’s 2020 American AI Initiative Report stated, “As the Federal Government increases its use of AI to improve the delivery of services to the American people, it is of paramount importance that this be done in a way that builds citizen trust. . . . Importantly, the Federal use of AI must always be done in a manner consistent with the Constitution and all applicable laws and government policies. Such development and use must show due respect for our Nation’s values, including privacy, civil rights, and civil liberties.”<sup>48</sup>

The 2024 House AI Taskforce Report found that “Improper use of AI can violate laws and deprive Americans of our most important rights.”<sup>49</sup> Based on these findings the report recommended:

---

<sup>44</sup> See, Joanna Bryson, Virginia Dignum et.al, *From fear to action: AI governance and opportunities for all*, (2023), Front. Comput. Sci. 5:1210421, doi:10.3389/fcomp.2023.1210421

<sup>45</sup> CAIDP, *Statement to U.S. Senate Judiciary Committee hearing on “Oversight of AI: Insiders’ Perspective”* (Sept. 17, 2024)

<sup>46</sup> A Safe Harbor for Independent AI Evaluation, <https://sites.mit.edu/ai-safe-harbor/>

<sup>47</sup> Merve Hickok, Evanna Hu, *Don’t Let Governments Buy AI Systems That Ignore Human Rights*, Issues in Science and Technology 40, no. 3, 37–41 (Spring 2024), <https://doi.org/10.58875/ACOG7116>

<sup>48</sup> American AI Initiative Report, pg. 25

<sup>49</sup> Bipartisan House Task Force Report on Artificial Intelligence, 118<sup>th</sup> Congress, pg. xii (December 2024), <https://www.speaker.gov/wp-content/uploads/2024/12/AI-Task-Force-Report-FINAL.pdf>



- “Have humans in the loop to actively identify and remedy potential flaws when AI is used in highly consequential decision-making.
- Agencies must understand and protect against using AI in discriminatory decision-making.
- Empower sectoral regulators with the tools and expertise to address AI-related risks in their domains.
- Explore transparency for users affected by decisions made using AI.”
- Support standards and technical evaluations to mitigate flawed decision-making involving AI systems.”

OSTP should consider the workforce impacts of AI systems including displacement, increased surveillance which disproportionately impact low-wage workers, and the upskilling/re-skilling needs for an AI-ready workforce.<sup>50</sup> Unregulated deployment of AI technologies can exacerbate exploitative and discriminatory business practices and can have severe detrimental impacts on the American’s rights, economy, and workforce.<sup>51</sup>

OSTP Director Michael Kratsios previously stated, “We can advance emerging technology in a way that reflects our values of freedom, human rights and respect for human dignity.”<sup>52</sup> The 2020 Executive Order stated “Agencies are encouraged to continue to use AI, when appropriate, to benefit the American people. The ongoing adoption and acceptance of AI will depend significantly on public trust.”<sup>53</sup> The OSTP should lead interagency coordination to ensure that systems are not deployed in the federal government that violate rights or threaten public safety. Additionally, the OSTP should incentivize the private sector to develop and deploy AI which supports the public trust priority of the government.<sup>54</sup>

## 8) Understandable/Explainable AI

The US AI Action Plan should ensure that the operations and outcomes of AI systems are sufficiently understandable by subject matter experts and users.<sup>55</sup> AI systems should be created to

---

<sup>50</sup> Stuart Russell, Daniel Dewey, Max Tegmark, *Research Priorities for Robust and Beneficial Artificial Intelligence*, <https://arxiv.org/abs/1602.03506>

<sup>51</sup> CAIDP, *Statement to U.S. Senate Committee on Health, Education, Labor, and Pensions hearing on “Reading the Room: Preparing Worker’s for AI”*, (Sept. 25, 2024).

<sup>52</sup> Michael Kratsios, Office of Science & Technology Policy, *AI That Reflects American Values - We don’t have to decide between freedom and technology* (Jan. 8, 2020), <https://trumpwhitehouse.archives.gov/articles/ai-that-reflects-american-values/>

<sup>53</sup> Section 1

<sup>54</sup> EO 13960, Section 2 (b)

<sup>55</sup> L. Floridi, J. Cowls, M. Beltrametti, M. et al., *AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*. *Minds & Machines* 28, 689–707, 2018, <https://doi.org/10.1007/s11023-018-9482-5>

give individuals the opportunity to contest adverse decisions relating to algorithmic bias and to access the logic, data and functions that led to the outcome.<sup>56</sup>

The International AI Safety Report on risk-management techniques for general purpose AI states “Developers still understand little about how their general-purpose AI models operate. This lack of understanding makes it more difficult both to predict behavioural issues and to explain and resolve known issues once they are observed.”<sup>57</sup> The AI Safety Report also emphasizes “an information gap between what AI companies know about their AI systems and what governments and non-industry researchers know.”<sup>58</sup>

The American AI Initiative report explicitly stated, “Priority areas of AI R&D emphasize the development of explainability mechanisms that help human users understand reasons for AI outputs, along with methods to test, evaluate, verify, and validate their performance.”<sup>59</sup> We urge the OSTP to carry forward this priority and disincentivize black box AI systems, especially in benefits administration.<sup>60</sup>

## 9) Traceable AI

The US AI Action Plan should ensure that AI systems are well documented, and that inputs and outputs of particular AI systems are traceable. Traceability is a common risk management practice in other high-stakes safety industries like bio-safety, nuclear safety etc.<sup>61</sup> As Gregory Falco and Ben Shneiderman explained:

Traceability or requirements tracing, which in turn facilitates transparency and explainability. An audit trail for highly automated system operations, for instance, could provide high-fidelity data to improve traceability, and, by extension, enable accountability. Analysts could thereby either identify risks using real-time monitoring and analysis, or provide post-event insight into the context surrounding the accident. If data regarding highly automated system accidents, including near misses, were stored in a transparent, publicly available data repository, they would be a valuable source for researchers, authorities and developers.<sup>62</sup>

---

<sup>56</sup> CAIDP, *Comments to OSTP on the National Priorities for AI* (Jul. 7, 2023), <https://www.caidp.org/app/download/8466307763/CAIDP-Statement-OSTP-07072023.pdf>

<sup>57</sup> Pg. 21

<sup>58</sup> Pg. 22

<sup>59</sup> Pg. 6

<sup>60</sup> See, Frank Pasquale, Gianclaudio Malgieri, *Generative AI, Explainability, and Score-Based Natural Language Processing in Benefits Administration*, *Journal of Cross-Disciplinary Research in Computational Law*, (forthcoming, 2024), <https://ssrn.com/abstract=4826707>

<sup>61</sup> International AI Safety Report, pg. 165

<sup>62</sup> Gregory Falco, Ben Shneiderman, *et. al.*, *Governing AI safety through independent audits*, *Nature Machine Intelligence*, 3, 566–571, 2021, <https://doi.org/10.1038/s42256-021-00370-7>

Such practices are well established in aviation with, for example, flight data recorders. Traceability is vital to US national security as well because it enables proactive threat monitoring, assessments, and response.

The American AI Initiative and Federal Data Strategy required agencies to “improve their data and model inventory documentation to enable discovery and usability, prioritizing access and quality improvements to support AI research, development, and testing.”<sup>63</sup> AI actors should ensure traceability, including in relation to datasets, processes and decisions made during the AI system lifecycle, to enable analysis of the AI system’s outcomes and responses to inquiry, appropriate to the context and consistent with the state of art.<sup>64</sup>

### 10) Algorithmic Fairness

The US AI Action Plan should promote AI transparency in content-moderation and all AI systems impacting the lives of Americans. One stated purpose of the January 2025 AI Executive Order is to develop AI systems “free from ideological bias or engineered social agendas.”<sup>65</sup> The Universal Guidelines for AI, widely endorsed by AI experts, states that: “institutions must ensure that AI systems do not reflect unfair bias or make impermissible discriminatory decisions. The Fairness Obligation recognizes that all automated systems make decisions that reflect bias, but such decisions should not be normatively unfair or impermissible.”<sup>66</sup>

The principle of algorithmic fairness is not simply a matter of the output or result of the system but the requirement that the functioning must be procedurally fair and transparent. As CAIDP founder Marc Rotenberg explained:

Algorithmic transparency is the basis of machine accountability. Credit determinations, employment assessments, educational tracking, as well as decisions about government benefits, border crossings, communications surveillance and even inspections in sports stadiums increasingly rely on black box techniques that produce results that are unaccountable, opaque, and often unfair.

---

<sup>63</sup> American AI Initiative Report, Pg. 10

<sup>64</sup> Department of Commerce, National Telecommunications and Information Administration (NTIA), *U.S. Joins with OECD in Adopting Global AI Principles*, May 22, 2019, <https://www.ntia.doc.gov/blog/2019/us-joins-oecd-adopting-global-ai-principles>

<sup>65</sup> White House Fact-Sheet 2025

<sup>66</sup> CAIDP, *Comments to OSTP on the National Artificial Intelligence Research and Development Strategic Plan*, Mar. 4, 2022, pg.3, <https://www.nitrd.gov/rfi/ai/2022/87-FR-5876/NAIRDSP-RFI-2022-combined.pdf>

Even the organizations that rely on these methods often do not fully understand their impact or their weaknesses.<sup>67</sup>

Algorithmic fairness is even more crucial for advanced AI systems such as large language models or generative AI systems where developers have already expressed an inability to address bias in their constructed datasets.<sup>68</sup> For example, OpenAI's GPT-4 system card states, "We found that the model has the potential to reinforce and reproduce specific biases and worldviews, including harmful stereotypical and demeaning associations for certain marginalized groups."<sup>69</sup> Subsequently, there have been several reports of ChatGPT's bias towards women. Speaking on the problems of machine learning systems, AI experts state, "OpenAI mitigates biases using reinforcement learning and instruction fine-tuning. But these methods can only correct the model's explicit biases, that is, what it actually outputs. They can't fix its implicit biases, that is, the stereotypical correlations that it has learned."<sup>70</sup>

General-purpose AI systems should be subject to validation for fairness and transparency. The way to detect bias in the medical sphere is not the same as detecting bias in the criminal justice system.<sup>71</sup> Unfair and opaque AI systems will diminish public trust and thus slow down AI adoption. In fact recognizing the significance of this issue the first Trump Administration OMB Memo states As the OMB has said, When considering regulations or non-regulatory approaches related to AI applications, agencies should consider, in accordance with law, issues of fairness and nondiscrimination with respect to outcomes and decisions produced by the AI application at issue."<sup>72</sup> With the proliferation of various AI models being developed across the world and released into global marketplaces, it is important that the US AI Action Plan prioritizes ex-ante assessments of AI systems prior to release so that models with harmful biases are not marketed and monetized at the expense of consumer protection and public trust.

---

<sup>67</sup> Marc Rotenberg, *Artificial Intelligence and the Right to Algorithmic Transparency*, in *The Cambridge Handbook of Information Technology, Life Sciences and Human Rights*, 2022, pg. 153 – 165, <https://doi.org/10.1017/9781108775038.015>

<sup>68</sup> Aylin Caliskan, Joanna Bryson, Arvind Narayanan, *Semantics derived automatically from language corpora contain human-like biases*, *Science*, 356.6334 (2017): 183-186. <http://opus.bath.ac.uk/55288/>; CAIDP, *Cases – In the matter of OpenAI and ChatGPT*, Federal Trade Commission, <https://www.caidp.org/cases/openai/>;

<sup>69</sup> CAIDP, *Complaint to the Federal Trade Commission in re OpenAI and ChatGPT*, Mar. 30, 2023, pg. 11, <https://www.caidp.org/app/download/8450269463/CAIDP-FTC-Complaint-OpenAI-GPT-033023.pdf>

<sup>70</sup> Arvind Naryanan and Sayash Kapoor, *AI Snake Oil: Quantifying ChatGPT's gender bias*, Apr. 26, 2023, <https://www.aisnakeoil.com/p/quantifying-chatgpts-gender-bias>

<sup>71</sup> Doaa Abu-Elyounes, *Contextual Fairness: A Legal and Policy Analysis of Algorithmic Fairness*, *Journal of Law, Technology, & Policy*, Vol. 2020, <https://illinoisjltlp.com/file/23/Elyounes.pdf>

<sup>72</sup> Office of Management and Budget, *Memorandum for the Heads of Executive Departments and Agencies – Guidance for the Regulation of Artificial Intelligence Applications*, M-21-06, Pg 6

The US AI Action Plan should favor robust mechanisms for algorithmic transparency that provide access to the logic, factors, and data that provide the basis for the outputs.<sup>73</sup> Companies using AI impacting public safety or American values should disclose the use of such techniques, and developers of foundation models should be required to implement specific safeguards.<sup>74</sup>

Fairness and transparency for social media algorithms are critical to ensure that tech companies are not infringing users' free speech rights.<sup>75</sup> The US AI Action Plan should encourage investment and innovation in advanced AI models that also distinguish between human and AI-generated content. This is akin to building an 'AI immune system', which can detect and flag synthetic media by identifying subtle inconsistencies or anomalies that are characteristic of machine-generated outputs.<sup>76</sup>

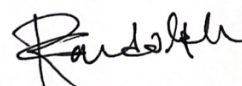
Thank you for the consideration of our views.



Marc Rotenberg  
Executive Director



Merve Hickok  
President



Christabel Randolph  
Associate Director



Maame Barnie Amoah  
Research Assistant

Leonor Ribeiro  
Research Assistant

*"This document is approved for public dissemination. The document contains no business-proprietary or confidential information. Document contents may be reused by the government in developing the AI Action Plan and associated documents without attribution."*

<sup>73</sup> Marc Rotenberg, *Artificial Intelligence and the Right to Algorithmic Transparency*, in *The Cambridge Handbook of Information Technology, Life Sciences and Human Rights* (Cambridge 2022)

<sup>74</sup> UNESCO, *Recommendation on the Ethics of Artificial Intelligence*, (Nov. 23, 2021) <https://en.unesco.org/about-us/legal-affairs/recommendation-ethics-artificial-intelligence>

<sup>75</sup> See, Marc Rotenberg, *U.S. Supreme Court: NetChoice Cases Explore AI and the First Amendment*, Case Note, *Journal of AI Law and Regulation*, Issue 2, 2024, <https://doi.org/10.21552/aire/2024/2/13>;

<sup>76</sup> CAIDP, *Comment to the OSTP and PCAST on Generative Artificial Intelligence*, Aug. 3, 2023, <https://s899a9742c3d83292.jimcontent.com/download/version/1693227547/module/8470565263/name/CAIDP-Statement-PCAST-GenAI-08032023..pdf>; The College Fix, *Conservative scholars tackle AI with 'Technology Agenda' for the future*, Mar. 5, 2025, <https://www.thecollegefix.com/conservative-scholars-tackle-ai-with-technology-agenda-for-the-future/>